Final Report on ONR grant N000141010221, Information Fusion of Human Activity, Social Networks, and Geography Using Fast Compressive Sensing

Personnel:

List of PIs: Andrea Bertozzi, P. Jeffrey Brantingham, S. J. Osher (UCLA); George Tita (UC Irvine)

Postdocs: (b) (6) [redacted]

Ph. D. students: (b) (6) [redacted]

Undergraduates: (b) (6) [redacted]

Summary

Research Objectives: We considered complex compressive sensing problems of interest to the Navy. The goal of this project was to develop new models and algorithms for human activity with both geographical and social network components. With datasets both from domestic crime and insurgent activity, we identified patterns of activity and ways to exploit those patterns for information recovery. We also considered some applied research problems involving imaging through turbulence.

Research Results: Results of this research fit into several different categories: algorithms for turbulence, temporal clustering models involving self-exciting point processes, spatial and social network clustering models using ideas from compressive sensing, density estimation, and forward models for human activity.

Transitions: We transitioned turbulence imaging software to China Lake; we transitioned self-exciting point process models to the company PredPol and to the Los Angeles Police Department. Pred Pol is currently deploying this software in over 20 cities worldwide. Data clustering software was co-developed with Arjuna Flenner at China Lake and is currently being implemented in Navy Applications.

List of publications and preprints:

1. Huiyi Hu, Thomas Laurent, Mason A. Porter, Andrea L. Bertozzi, A Method Based on Total Variation for Network Modularity Optimization using the MBO Scheme accepted in *SIAM J. Appl. Math.*, 2013.
2. Cristina Garcia-Cardona, Ekaterina Merkurjev, Andrea L. Bertozzi, Arjuna Flenner and Allon G. Percus, Fast Multiclass Segmentation Using Diffuse Interface Methods on Graphs, submitted 2013.
3. Yifei Lou, Sung Ha Kang, Stefano Soatto, and Andrea L. Bertozzi, Video Stabilization of Atmospheric Turbulence Distortion, *Inverse Problems in*

Imaging, 7(3) pp. 839-861, 2013.

4. Mario Micheli, Yifei Lou, Stefano Soatto, and Andrea L. Bertozzi, A linear systems approach to imaging through turbulence, *JMIV*, online first July 2013.
5. Rachel A. Hegemann, Erik A. Lewis, and Andrea L. Bertozzi, An "Estimate & Score Algorithm" for simultaneous parameter estimation and reconstruction of missing data on social networks *Security Informatics*, 2:1, 2013.
6. Yves van Gennip, Blake Hunter, Raymond Ahn, Peter Elliott, Kyle Luh, Megan Halvorson, Shannon Reid, Matt Valasik, James Wo, George E. Tita, Andrea L. Bertozzi, P. Jeffrey Brantingham, Community detection using spectral clustering on sparse geosocial data, *SIAP*, 73(1), pp. 67-83, 2013.
7. B. Hunter, Y. Lou, A.L. Bertozzi, A spectral graph based approach to analyze hyperspectral data, *IEEE Applied Imagery Pattern Recognition*, Washington DC, 2012.
8. J.Gilles, S.Osher, "Fried Deconvolution", *SPIE Defense, Security and Sensing Conference*, Baltimore, US, April 2012.
9. H. Hu, Y. van Gennip, B. Hunter, A.L. Bertozzi, M.A. Porter, Multislice Modularity Optimization in Community Detection and Image Segmentation, *IEEE International Conference on Data Mining (Brussels), ICDM'12*, 2012.
10. Tijana Kostic and Andrea Bertozzi, Statistical Density Estimation using Threshold Dynamics for Geometric Motion, *J. Sci. Comp*., 54(2-3), pp. 513-530, 2013.
11. Yves van Gennip and Andrea L. Bertozzi, Gamma-convergence of graph Ginzburg-Landau functionals, *Advances in Differential Equations*, 17(11-12), 2012, pp. 1115-1180.
12. Laura M. Smith, Andrea L. Bertozzi, P. Jeffrey Brantingham, George E. Tita, and Matthew Valasik, Adaptation of an Ecological Territorial Model to Street Gang Spatial Patterns in Los Angeles *Discrete and Continuous Dynamical Systems A*, 32(9), pp. 3223 - 3244, 2012.
13. Alexey Stomakhin, Martin B. Short, and Andrea L. Bertozzi, Reconstruction of Missing Data in Social Networks Based on Temporal Patterns of Interactions, *Inverse Problems*, 27(11), 115013, 2011.
14. E. Lewis, G. Mohler, P. J. Brantingham, and A. L. Bertozzi, Self-Exciting Point Process Models of Civilian Deaths in Iraq, *Security Journal*, 25, 244-264, 2012.
15. R. A. Hegemann, L. M. Smith, A. Barbaro, A. L. Bertozzi, S. Reid, and G. E. Tita, Geographical influences of an emerging network of gang rivalries, *Physica A,* Volume 390, Issues 21-22, 15 October 2011, Pages 3894-3914.

Detailed discussion of results:

Point process models: We examined self-exciting point processing models as a way of describing repeat victimization in time for both domestic crime and insurgent activity. This work builds on earlier work by some of the PIs using such models to describe residential burglaries. For insurgent activity we considered data from the Iraq Body Count (civilian casualties) between 2003 and 2007. The rate of violent events is partitioned into a background rate and a foreground self-exciting component. Background rates are assumed to change on relatively long time-scales. Foreground self-excitation, in which events trigger an increase in the rate of violence, is assumed to be short-lived. Our results indicate that self-excitation makes up as much as 37 – 50 per cent of all violent events and that self-excitation lasts at most between two and six weeks, depending upon the district in question. Appropriate security responses may benefit from taking these different time-scales of violence into consideration. This work was published in *Security Journal.* We were contacted by the Army to further develop these results into a field application. However they were unable to find funding to further pursue that line of inquiry.

We also considered self-exciting point processes on a social network as a model for gang crimes in the Hollenbeck Division of LAPD. Earlier analysis of gang data showed that such a model well-described temporal patterns of activities between rival street gangs. The big challenge for law enforcement is to solve the numerous crimes that go unsolved in this area – hundreds each year. The metric presented by LAPD as a goal for the mathematical algorithms was to identify the three gangs most likely to have committed a crime. This information was believed to aid police investigations and result in a high probability of solving the crime. Our algorithm was able to improve the statistics from 50% chance of identifying the offending gang in the top three (from random guessing based on the known rivalry network) to about 80% chance by using temporal patterns of information. The method is based on a variational model involving assignments of weights to rivalry pairs for each unsolved crime and optimizing a functional that includes statistics of the self-exciting point process. This work was published in *Inverse Problems.* In a second paper we considered the problem of simultaneous parameter estimation and filling in missing data. This work was closer to the real world problems than the first work in which parameters for the model were assumed to be known. The second work involved additional algorithm development resulting in an iterative method that estimates both parameters and unsolved crimes - it resulted in a paper published in *Security Informatics.*

Social Network Clustering:

We propose an agent-based model to simulate the creation of street gang rivalries. The movement dynamics of agents are coupled to an evolving network of gang rivalries, which is determined by previous interactions among agents in the system. Basic gang data, geographic information, and behavioral dynamics suggested by the criminology literature are integrated into the model. The major highways, rivers, and the locations of gangs' centers of activity influence the agents' motion. We use the Hollenbeck Division of LAPD as our test problem. We apply common metrics from graph theory to analyze the model, comparing networks produced by our simulations and an instance of a Geographical Threshold Graph to the existing network from the criminology literature. This work was published in *Physica A.*

We further explore the role of geography in social network development by adapting a territorial model from ecology due to Moorcroft et al. Coyotes and wolves organize themselves around a den site and mark their territory to distinguish their claimed region. We modify this ecological approach to simulate spatial gang dynamics in the Hollenbeck policing division of eastern Los Angeles. We incorporate important geographical features from the region that would inhibit movement, such as rivers and freeways. From the gang and marking densities created by this method, we create a rivalry network from overlapping territories and compare the graph to both the observed network and those constructed through other methods. Data on the locations of where gang members have been observed is then used to analyze the densities created by the model. Results from the model are quantitatively consistent with locations of known violence in this area of Los Angeles. This work was published in *DCDS* (Smith et al).

In a recent *SIAP* article (van Gennip et al), we identify social communities among gang members in the Hollenbeck policing district in Los Angeles, based on sparse observations of a combination of social interactions and geographic locations of the individuals. This information, coming from Los Angeles Police Department (LAPD) Field Interview cards, is used to construct a similarity graph for the individuals. We use spectral clustering to identify clusters in the graph, corresponding to communities in Hollenbeck, and compare these with the LAPD's knowledge of the individuals' gang membership. We analyze different ways of encoding the geosocial information using a graph structure and the influence on the resulting clustering. We study the robustness of this technique with respect to noisy and incomplete data, and provide suggestions about the relative importance of quantity versus quality of collected data.

The spectral clustering method is computationally tractable but may not produce correct results when the data is high dimensional and has complex geometric structures. Another method that is computationally difficult but produces good results is Modularity Optimization. We develop new algorithms to compute network modularity by making a connection to problems in graph cuts. We reformulate modularity optimization as a minimization problem of an energy functional that consists of a total variation term and a balance term on graphs. By employing numerical techniques from image processing and compressive sensing such as convex splitting and the Merriman-Bence-Osher (MBO) scheme we develop a variational algorithm for the minimization problem. We test the algorithm using both synthetic benchmark networks and real data from machine learning problems. This work has been accepted for publication in *SIAP.*

Imaging through turbulence:

In this paper we address the problem of recovering an image from a sequence of distorted versions of it, where the distortion is caused by what is commonly referred to as ground-level turbulence. In mathematical terms, such distortion can be described as the cumulative effect of a blurring kernel and a time-dependent deformation of the image domain.

(a) We introduce a statistical dynamic model for the generation of turbulence based on linear dynamical systems (LDS). We expand the model to include the unknown image as part of the unobserved state and apply Kalman filtering to estimate such state. This operation yields a blurry image where the blurring kernel is effectively isoplanatic. Applying blind nonlocal Total Variation (NL-TV) deconvolution yields a sharp final result. This work was recently published in *JMIV* (Micheli et al).

(b) In a second work, we propose a method combining Sobolev gradient and Laplacian to stabilize the video sequence, and a latent image is further found utilizing the ``lucky region" method. The video sequence is stabilized while keeping sharp details, and the latent image shows more consistent straight edges. We analyze the well-posedness for the stabilizing PDE and the linear stability of the numerical scheme. This work was recently published in *Inverse Problems in Imaging* (Lou et al) and the software has been transitioned to China Lake.

(c) In a third work, we presented a new approach to deblur the effect of atmospheric turbulence in the case of long range imaging. Our method is based on a beautiful analytic formula due to Fried- the Fried kernel, which was never before used in atmospheric deblurring algorithms. This transfer function is used together with a framelet based deconvolution method which used split Bregman for very rapid convergence. The only important parameter requiring special measurements is the refractive index. We use total variation to devise a method which gives a good estimate of the parameter from the input blurred image. The algorithms are state-of-the-art in performance, and easy to implement. We applied them to simulated and real images. This work was published in the *SPIE Conf. on Defense Security and Sensing 2012.*

General clustering of high dimensional data:

There are currently several communities working on algorithms for classification of high dimensional data. We develop a class of variational algorithms that combine recent ideas from spectral methods on graphs with nonlinear edge/region detection methods traditionally used in the PDE-based imaging community. The algorithms are based on the Ginzburg–Landau functional which has classical PDE connections to total variation minimization. Convex-splitting algorithms allow us to quickly find minimizers of the proposed model and take advantage of fast spectral solvers of linear graph-theoretic problems. We present diverse computational examples involving both basic clustering and semisupervised learning for different applications. Case studies include feature identification in images, segmentation in social networks, and segmentation of shapes in high dimensional datasets. This work is joint with Arjuna Flenner from China Lake. The GL-binary classifier algorithm was published in *MMS*. A second paper on multi-class classification is currently under review. The method was also adapted to Modularity Optimization as discussed above. Rigorous theory for these methods was proved in some recent papers include the *ADE* paper by van Gennip and the PI in 2012. These algorithms are currently being transitioned to China Lake. We are also in discussion with SPAWAR regarding the testing of these methods.